

EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly
journal
Editor: Gene V Glass
College of Education
Arizona State University

Copyright is retained by the first or sole author, who grants right of first publication to the **EDUCATION POLICY ANALYSIS ARCHIVES. EPAA** is a project of the [Education Policy Studies Laboratory](#).

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the [ERIC Clearinghouse on Assessment and Evaluation](#) and are permanently archived in *Resources in Education*.

Volume 12 Number 16

April 10, 2004

ISSN 1068-2341

Education and Alternate Assessment for Students with Significant Cognitive Disabilities: Implications for Educators

Mary C. Zatta
Perkins School for the Blind

Diana C. Pullin
Boston College

Citation: Zatta, M., Pullin, D., (2004, April 10). Education and alternate assessment for students with significant cognitive disabilities: Implications for educators. *Education Policy Analysis Archives*, 12(16). Retrieved [Date] from <http://epaa.asu.edu/epaa/v12n16/>.

Abstract

State and federal mandates for education reform call for increased accountability and the inclusion of students with disabilities in all accountability efforts. In the rush to implement high-stakes education reforms, particularly those involving tests or assessments, the particular needs of students with severe cognitive disabilities are only now being addressed by policymakers and educators. For students with significant cognitive disabilities, implementation of alternate approaches to education accountability is increasing. At the same time, the challenges associated with successfully implementing alternate assessment programs are becoming more obvious. This paper describes some of the ways in which alternate assessment as

part of standards-based education reform may impact students with significant cognitive disabilities. It provides an overview of state efforts to implement alternate assessments for students with significant cognitive disabilities, followed by an example of how one state has begun to implement alternate assessment through the Massachusetts Alternate Assessment (MCAS-Alt/ Massachusetts Comprehensive Assessment System Alternate). It reviews issues educators in all states will face in the participation of students with significant disabilities in alternate assessment programs, the content and form of alternate assessments, the validity and reliability of the assessments, and the role of teachers in the implementation of alternate assessment programs.

Education reform has become one of the paramount public policy issues in the nation. As policymakers and educators rush to rectify the many perceived shortcomings of our educational system by requiring more accountability, it is increasingly clear that many reforms have not, in fact, fully taken into consideration the particular needs of students with significant cognitive disabilities. For these students, the implementation of alternate approaches to education accountability is increasing. At the same time, there is limited guidance from research on how to appropriately implement alternate assessment and local educators have limited preparation in alternate assessment practices. This paper describes some of the ways in which alternate assessment as part of standards-based education reform may impact students with significant cognitive disabilities. It provides an overview of state efforts to implement alternate assessments for students with significant cognitive disabilities, followed by an example of how one state has begun to implement alternate assessment through the Massachusetts Alternate Assessment (MCAS-Alt / Massachusetts Comprehensive Assessment System Alternate). Then, it reviews some of the potential issues researchers and educators in all states will face in the participation of students with significant disabilities in alternate assessment programs, the content and form of alternate assessments, the validity and reliability of the assessments, and the role of teachers in the implementation of alternate assessment programs.

Standards-Based Education Reform: Mandates for Accountability

The current wave of education reform initiatives extends back to the mid-1980s, when national calls for dramatic change began to draw considerable public attention to the quality of schools and the need for increased accountability for educational outcomes (National Commission on Education, 1983). Eventually, a movement calling for systemic reform of the nation's schools was born. This initiative focused upon an effort to impact all components of the educational process in an effort to achieve pervasive and meaningful change. The dissatisfaction with American education led to a shift in focus "from the process of education to the outcomes of the educational process" (Geenen, Thurlow, & Ysseldyke, 1995, p. 2). By the mid-1990s, the states began to establish educational standards and outcomes, often relying heavily upon the use of high-stakes tests to both define and measure educational progress. The U.S.

Congress declared the importance of embracing the goal of ensuring that “all children can learn and achieve to high standards” and set out incentives to insure that all states pursued this goal (Goals 2000: Educate America Act of 1994 (P.L. 103-227). At about the same time, and for the first time, Congress declared in both its special education laws and its general legal requirements for elementary and secondary education that high standards and accountability should apply to all students, including students with disabilities (U.S. P.L. 103-227, Section 3(1), 1994; Title I of the Improving America’s Schools Act (IASA) of 1994; Individuals with Disabilities Education Act (IDEA) of 1997). The 1997 amendments to the IDEA mandated the alignment of general and special education reform efforts (Guy, Shin, Lee, & Thurlow, 1999).

The IDEA’97 requires that children with disabilities be included in general state and district-wide assessment programs. The mandate underscores that accommodations be provided for students with disabilities to ensure appropriate participation in the assessment. Further, for those students with significant disabilities, IDEA '97 requires that each state provide an alternate assessment for those children who cannot participate in the standard State and district-wide assessment programs. Finally, the law places the responsibility upon each state for developing the participation guidelines and gives the IEP team responsibility for making determinations on the participation of each student in state assessment programs based on the state guidelines.

In the No Child Left Behind Act of 2001 (NCLBA) (P.L. 107-110), Congress reaffirmed and expanded its commitment to standards-based education reform. The new law requires annual testing of students in grades three through eight, calls for determinations whether schools are making “adequate yearly progress” in meeting academic standards, and encourages greater accountability for educational progress, including the use of sanctions and rewards. The NCLBA also addresses the participation of students with disabilities in these programs. In assessing adequate yearly progress, it calls for participation of no less than 95% of students with disabilities in either regular assessment or alternate assessment programs, reasonable adaptations and accommodations for students with disabilities, the use of valid and reliable measures for students with disabilities, disaggregated accountability reporting to focus on outcomes for students with disabilities, and meaningful reporting to parents of individual student results.

The essential components of all these recent reform mandates rest upon the use of content standards, performance assessments, and accountability. Initially, content standards were the main political tools of standards-based reform: “They define the breadth and depth of valued knowledge that students are expected to learn, and they are intended to reduce the curriculum disparities existing across schools and school districts” (McDonnell et al., 1997, p. 114; see also Ysseldyke, Thurlow, & Shriner, 1994). Performance assessment, however, became the mechanism for ensuring accountability in meeting academic content standards.

Accountability is central to standards-based reform and takes two forms: student accountability (assigns responsibility to the student) and system accountability (assigns responsibility to the educational system or individuals

within that system). "System accountability is designed to improve educational programs whereas student accountability is designed to motivate students to do their best" (National Center on Educational Outcomes, 2001). System accountability, defined as "a system activity designed to assure those inside and outside the educational system that schools are moving in desired directions" (p. 2), is most often measured by large-scale standardized tests (Ysseldyke, Olsen and Thurlow (1997). Student accountability is also most often attained through standardized tests and is many times linked to high school graduation or grade-to-grade promotion requirements. According to the National Center for Educational Outcomes, "all states have some type of system accountability, but not all states have student accountability" (National Center on Educational Outcomes, 2001).

Until recently, there has generally been a dual system of accountability - one for general education and one for special education (Sebba, Thurlow, & Goertz, 2000). Indeed, some would argue that for students with disabilities there was no systemic accountability at all (McDonnell, et al., 1997). Now, there is a push for a unified educational accountability system based upon the realization that "accountability is only realized when *all* children, including students with disabilities, are considered in the planning, development, and implementation" (Erickson & Thurlow, 1997, p. 1).

For students with disabilities, inclusion in the general system for student and system accountability is intended to insure full participation in the content and performance standards of general education. These goals began to be addressed as students with disabilities were included in state and local large-scale testing programs. For some, this participation required some accommodations or modifications to allow participation. However, for the much smaller population of students with significant disabilities, participation in large-scale assessment programs, even with accommodations or modifications is not appropriate. For the population of students with significant disabilities, alternate assessment systems are now being implemented to address the mandates for inclusion of *all* students in assessment and accountability programs. There are, however, significant challenges associated with the implementation of these alternate assessments.

Some of these challenges have been deliberated in the courts. Even the federal courts have become involved in struggles over alternate assessment. The courts have previously upheld the right of states and local districts to make high-stakes decisions, such as the award of a high school diploma contingent upon student test performance (*Debra P.v. Turlington*, 1981; *Brookhart v. Illinois State Board of Education*, 1982; *Board of Education v. Ambach*, 1983). However, the courts also specified that tests used for these purposes had to be valid and based upon content that students had a fair opportunity to learn. They also required, for students with disabilities, that IEPs should create appropriate opportunities for students to prepare for tests. Recently, a federal district court mandated that the State of California must insure that students with learning disabilities, including those under both IEPs and Section 504 plans, must be provided alternate assessments if they are unable to access the general test due to a disability (*Chapman v. California Dept. of Ed.*, Feb 21, 2002).

Alternate Assessments – What are they?

For students with disabilities for whom participation in the general assessment program with accommodations is not appropriate, educators have turned to alternate assessment programs. The term "alternate assessment" has been defined by Ysseldyke, et al. (1997) as "any assessment that is a substitute way of gathering information on the performance and progress of students who do not participate in the typical state assessment used with the majority of students who attend school" (p. 2). Alternate assessment is seen as an "approach to enable the educational outcomes of students with the most significant disabilities to be included in school and district accountability measures" (Kleinert, Haig, Kearns, & Kennedy, 2000, p. 53; see also Coutinho and Malouf (1993). Thompson, Quenemoen, Thurlow, and Ysseldyke (2001) provide examples of alternate assessments explaining that "alternate assessments typically involve some variation of what is sometimes called performance-based assessment, authentic assessment, or 'alternative' assessment, or with a collection of these tools, portfolio assessment" (pp. 80-81). As portfolio assessments have become more common for performance assessment, they have become more systematic. Student accomplishments are systemically sampled or collected over a period of time to assess student growth and attainment in content areas (Baker, 1993). Portfolios are now being measured against predetermined scoring criteria (Thompson, et al., 2001).

Most states have adopted a portfolio assessment model as their method of alternate assessment for students with disabilities (Thompson et al., 2001,). Kentucky and Maryland have led the way in the implementation of alternate assessments. "Both of these states have used the idea of portfolio assessment as a means of gathering achievement information when students cannot participate in the general state assessments" (Rouse, Shriner, & Danielson, 2000, p. 89). However, the format for these assessments has been variable across the country (Thompson, et. al, 2001) and the research on implementation of these practices is thus far somewhat limited.

Carpenter, Ray & Bloom (1995) describe the benefit of portfolios in terms of their ability to provide concrete evidence of student work and progress toward annual goals and objectives. "The goal of these newer assessments is to more accurately depict what students can do, in more authentic or real-life contexts, and to focus classroom instruction on the development of problem-solving and higher-order thinking and writing skills" (Kleinert, Kennedy, & Kearns, 1999, p. 93). According to Thompson et al., (2001) and Choate & Evans (1992) there are numerous advantages to using a portfolio assessment model. These advantages include an increased ability for school districts to be accountable for all students, the ability to demonstrate student growth, an assessment process that is able to include all students on an individualized basis, a demonstration of student progress toward standards, and a "means of incorporating assessment and instruction relevant to functioning in the real world" (Choate & Evans, 1992, p. 9).

At the same time, there is growing recognition of some of the challenges posed by the use of portfolio assessments – difficulty with the implementation process,

scoring difficulty, problems with generalizability and comparability of results, and validity and reliability issues. Ysseldyke and Olsen (1997) warn that “there is little consensus on what constitutes a portfolio or how portfolios should be used in large-scale assessment” (p. 11). Another commentator (Maurer, 1996) speaks to the need for clarity regarding four specific issues about portfolio assessment: the purpose of portfolio assessment (why assess?), participation guidelines for portfolio assessment (who to assess?), alignment of the assessment with what is being taught (what to assess?), and the validity and reliability of the assessment (how to assess and score?). Each of these issues frames an essential set of questions for educators implementing alternate assessments.

Why Assess? The purpose or purposes of any assessment must be established at the outset. “Many of the technical issues presented by the conceptions of portfolio assessment in the literature could likely be resolved by clarifying the purpose of portfolios” (Nolet, 1992, p. 11). However, as Olsen (1998) noted in a review of state practices, “one of the common threads that runs through these documents is the need for states to establish a solid philosophical basis for alternate assessments before moving too far in to the details of development” (p. 1).

According to the National Center on Educational Outcomes, “the primary purpose for alternate assessments is to increase the capacity of large-scale accountability systems to create information about how a school, district, or state is doing in terms of overall student performance” (NCEO, 2000). In addition to these systemic accountability purposes, however, assessment results provide judgment or accountability information to the student and the parent (Maurer, 1996). These goals are not necessarily easily reconciled. For either systemic or student accountability, the basic premises of alternate assessments are the same. These assessments must be “designed to provide information relative to key performance indicators that represent the most essential features of the educational experience of students with disabilities” (Ysseldyke, Thurlow, Kozleski, & Reachly, 1998b, p. 14). Warlick (2000) discusses the importance of alignment of alternate assessments with each state’s general assessment: “the purpose of an alternate assessment should reasonably match, at a minimum, the purpose of the assessment for which it is an alternate” (p. 18).

In most programs, assessments including alternate assessments are seen as “a matter of school accountability more than student accountability” (Kleinert et al., 2000, p. 53). However, in many states and local school districts, there are also high-stakes accountability consequences for students, such as the determination of the type of exit credential a student may receive. And, even when high-stakes consequences may be limited for individual students, the availability of alternate assessment evidence can be expected to play a key role in such critical activities as the formulation or revision of IEPs. Multiple uses of alternate assessments may be significant particularly if there are high stakes involved. States must ensure that portfolio assessments measure what they are intended to measure and recognize that if they are being used for multiple purposes (e.g., student accountability and school accountability) that what they measure is consistent with the purposes of the assessment. Failure to meet

these requirements may have a significant impact on the validity of an assessment.

Who to Assess? States must develop specific guidelines regarding participation in alternate assessment. Consistent with IDEA '97 requirements, Warlick and Olson's (1998) report demonstrates that in all 12 states they surveyed, the IEP teams are called upon to make the decisions regarding whether students will participate in the general education test or the alternate assessment and to document justification for this decision in the IEP.

Appropriately, the task of specifying the criteria to be used in making these decisions are left up to the states. To date, numerous states have established participation guidelines. However, these guidelines are not consistent from state to state. Warlick and Olson's (1998) examined the practices in twelve states and found: 75% of the states use a curriculum focus criterion (i.e., unable to participate fully in the general curriculum, pursuit of functional or living skills oriented curriculum, etc.) in determining participation. Sixty-seven per cent of the states cited the student's need for "intensive individualized instruction in order to acquire, maintain, or generalize skills" as a criterion for alternate participation (Warlick & Olsen, 1998, p. 10). In some states (59%) older students are permitted to participate in an alternate assessment "only if they are unable to complete the regular diploma program even with program adaptations" (Warlick & Olsen, 1998, p. 10).

There is an overall concern about how to institute an alternate assessment process without once again creating a mechanism that promotes a dual educational system or other unintended consequences. One challenge focuses upon weighing the balance between the systemic and the individual accountability goals associated with a program. At the ground level, when individual IEP participants are making decisions about whether to include a student in the standard or the alternate assessment system, the primary consideration is probably the individual needs of the student. However, the influences associated with systemic accountability also must be in play. This is particularly true when there is a high-stakes impact on the school, the district, or even the individual educators who work with the student, as is the case in the growing number of states now seeking to measure teacher accountability on the basis of student assessment performance.

When the costs associated with systemic accountability are high, there might be a press to have larger numbers of students with disabilities included in alternate assessment as a means of preventing their scores from being factored in with the rest of the scores from the standard assessment. This practice might make overall system performance seem higher. But, "placing a large number of students with disabilities in an alternate assessment program.... could help perpetuate the separate system that has been a concern for many" (Warlick & Olsen, 1998, p. 3). And, certainly far from clear at this time is the impact of what might be viewed as a slight Congressional pull-back in the No Child Left Behind Act of 2001 from the previous commitment to participation of *all* children to allow only 95% participation in determining systemic accountability, or "adequate yearly progress".

What to Assess? The advocacy for curriculum standardization is a critical

component in the current reform movement. Yet, this point of view is not without problems. McIntyre (1992) saw the emphasis on curriculum standardization as a problem for special education in that it “would hinder individualization in special classes” (p. 7). Ysseldyke, Thurlow, & Geenen (1994) emphasize that the successful participation of students with disabilities is dependent on states developing “outcomes that are comprehensive and broad enough to be meaningful for all students” (p. 5). McDonnell, et al. (1997) also articulate a need for attention to the specific curricular needs of students with significant cognitive disabilities: “the degree to which a set of content standards is relevant to their valued educational outcomes and consistent with proven instructional practices will determine how successfully they will participate in standards-based reform” (p. 114).

In order to achieve comprehensive and broad outcomes without lowering standards, consensus must be reached among stakeholders on both standards and outcomes. McDonnell et al. (1997) describe the conflicts resulting from the differing assumptions of standards-based reform and special education and conclude that the successful participation of students with disabilities in standards-based reform will depend on the alignment between these assumptions. Standards-based reform has been built around a specific set of assumptions about curriculum and instruction, embodied in the content and performance standards that are central to the reforms. Special education, for its part, has been built around a set of assumptions about valued post-school outcomes, curricula, and instruction that reflect the diversity of students with disabilities and their educational needs. (McDonnell et al., 1997). Most parents and special educators agree that a functional curriculum approach is essential for students with severe cognitive disabilities. If the alternate assessment system can align with the general curriculum without precluding a simultaneous focus on functional life skills, how do we ensure that alternate assessment is appropriate and comprehensive and maintains a philosophical focus geared toward a unified education approach (i.e., no separate focus for special education)?

While there is a strong sentiment against the development of “separate standards” for the small percentage of the student population composed of students with significant disabilities (Ysseldyke & Thurlow, 1999), states have taken a range of approaches to alternate assessments. “Some states and districts focus very narrowly on specific academic standards, whereas others take a broader approach and include many functional or life skills within their standards for all students” (Thompson, et al., 2001, p. 22). One of the most prevalent concerns is about the “cost” of an academic focus for students who have participated in a more “functional” or “practical” program. Guy, et. al. (1999) addresses this concern “that students with disabilities may be merged into a system that has a heavy focus on academics, often to the exclusion of more applied and vocational kinds of skills, (the result of which) threatens what has been working for students with disabilities” (p. 78). Two leaders in the implementation of alternate assessment, the states of Kentucky and Maryland, while basing the assessment criteria on the core learning outcomes identified for all students, “clearly attempted to address the functional skill needs of students in their respective alternate assessments” (Kleinert et al., 2000, p. 57). A national study in 2000 reported this range of approaches by states:

- alternate assessments encompass general education standards in 28 states;
- alternate assessment in 7 states assess standards with an additional set of functional skills;
- two states have two alternate assessments - one that assesses general education standards at lower levels and one that assesses functional skills;
- alternate assessments in 3 states were developed based on functional skills and then linked back to state standards; and
- nine states based their alternate assessments on functional skills only with no alignment to state standards (Warlick, 2000).

The different possibilities open in selecting the content of alternate assessments present several challenges for educators. The possible tensions between student accountability purposes and systemic accountability purposes must be addressed. The extent to which inclusion for students with significant disabilities in the content standards of general education must be determined. States must continue to address these issues as they refine their standards-based reform efforts. States must continue to evaluate whether or not a dual education system is being perpetuated while at the same time examining the impact of content standards on students with significant disabilities.

How to Assess and Score? For any assessment, it is important to ensure that the resulting scores are accurate, reflect the information the assessment was intended to collect, and are meaningfully linked to teaching practice. In a report compiled by Quenemoen, Thompson and Thurlow (2003), comparing the assumptions and values embedded in the scoring criteria used in five states for their alternate assessments, discuss the importance of teachers having an understanding of “the stated and embedded scoring criteria” (p. 41). They caution states to keep in mind that “alternate assessments are a much more recent development than regular assessments (Quenemoen et al., 2003, p. 41)” and as such, advocate the necessity of ongoing debate and discussion regarding the underlying assumptions as they relate to students with significant cognitive disabilities and the impact of those assumptions on the scoring criteria.

The struggles involved in establishing reliable and valid test results are evidenced throughout the literature. Even without the particular complications associated with the alternate assessment of students with disabilities, one leading commentator on testing and assessment has noted that all types of performance assessment “present a number of validity problems not easily handled with traditional approaches and criteria for validity research” (Moss, 1992, p. 230). Other commentators have noted political problems associated with performance assessments: “If performance assessments are to gain any credibility with students, parents, and the community, they need to be reliable, valid, and generalizable. If we as a profession do not establish these traits, then performance assessments will, in time, come under the same type of attack that standardized tests receive today” (Maurer, 1996, p. 111).

Clearly, the concerns regarding validity and reliability have a critical impact for systemic and student accountability. Given the timelines involved in meeting federal mandates concerning both accountability and the inclusion of students with disabilities, the time required to establish reliability and validity has been short and the expertise on how to do so not widely available (Heaney & Pullin, 1998). The American Educational Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement in Education (NCME) have set the professional standards of practice for educational and psychological testing in their publication *Standards for Educational and Psychological Testing* (1999). While these requirements do not include extensive discussion of performance assessment issues, they do establish benchmarks for validity and reliability determinations that should be taken into account by educators implementing alternate assessment systems.

The *Test Standards* define validity as “the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests...the proposed interpretation refers to the construct or concepts the test is intended to measure” (AERA, APA, & NCME, 1999, p. 9). Caution must be taken when determining the types of evidence that might be incorporated into a portfolio or other performance assessment. “Important validity evidence can be obtained from an analysis of the relationship between a test’s content and the construct it is intended to measure” (AERA, APA, & NCME, 1999, p. 11). The evidence or work samples included in an assessment must support the construct or concepts being measured and they must be sufficient and relevant. Miller and Legg (1993) reference “eight criteria that need to be studied for serious validation of alternative assessments: intended and unintended consequences of test use, fairness, transfer and generalizability, cognitive complexity, content quality, content coverage, meaningfulness, and cost and efficiency” (p. 10).

The *Test Standards* (AERA, APA, & NCME, 1999) define reliability as “the consistency of such measurements when the testing procedure is repeated on a population of individuals or groups” (AERA, APA, & NCME, 1999, p. 25). After performance assessment results are collected, someone has to judge student responses and determine whether they meet the requisite educational standards. In scoring portfolio assessments, judges determine an individual’s score based on defined criteria or scoring rubrics. “Inter-rater reliability is also necessary in alternative assessments because the scoring procedures are usually subjective” (Miller & Legg, 1993, p. 11). Inter-rater scoring reliability plays an important role in establishing the validity of an assessment and is therefore, subject to rigorous technical requirements. “In such cases relevant validity evidence includes the extent to which the processes of the observers or judges are consistent with the intended interpretation of scores” (AERA, APA, & NCME, 1999, p. 13). Establishing the reliability of such judgments on a large-scale assessment program has already been identified as a significant challenge (Shepard, 1992); many more issues arise when alternate assessments are being administered.

Vermont was one of the first states to use portfolio assessments on a large-scale basis for all students, including those with disabilities. Koretz,

McCaffrey, Klein, Bell, & Stecher (1993) evaluated the 1992 Vermont Portfolio Assessment program and found disappointing reliability coefficients. In Kentucky, another state on line early with these assessments, there was an early finding that “there remains much work to be done around issues of reliability of scoring procedures” (Elliott, 1997 p. 106; see also Koretz & Hamilton, 2000). Sailor (1997) found that “the Kentucky experiment with Alternate Portfolios is plagued with predictable problems of reliability of judgment across independent scorers” (p.103). In Kentucky, portfolios were scored initially by the teachers administering them. This led to a concern regarding subjectivity, especially because Kentucky’s statewide assessment system was a high stakes system. Schools in Kentucky are subject to rewards and sanctions based on the assessment scores. When an assessment system is a high stakes system, it is subject to even greater scrutiny regarding validity and reliability because of the ultimate “cost”, or consequences, of the assessment results. The inter-rater reliability in Kentucky has shown a substantial increase since the mandate that every alternate portfolio “be blindly and separately scored by two trained scorers and that all disagreements be reconciled through a third, state-level scoring” (Kleinert, et al., 2000, p. 60).

Another significant issue regarding the validity and reliability of the alternate assessment are issues of whether or not the portfolio is a reflection of the student’s work or the teacher’s abilities. A statewide teacher survey conducted by Kleinert, et al., (1999) noted a concern regarding “the extent to which the alternate assessment was more of a teacher assessment than a student assessment” (p. 93). In portfolio assessment the resulting product to be judged for accountability purposes is a compilation of the student’s work. Students with significant disabilities are typically reliant on teachers to assemble their portfolio. The question arises as to the degree the resulting product is more reflective of the teacher’s expertise in assembling a portfolio that meets the requirements of the scoring rubric than the capabilities of the student. Is the resulting score a measure of the student’s ability and achievement or the teacher’s ability to assemble a portfolio to meet the specifications of the assessment? In the Kentucky statewide teacher survey, teachers’ comments indicated a concern that “teacher creativity/work is a greater factor in determining the ultimate score than is student learning” (Kleinert et al., 1999, p. 98).

The mandates for available and persuasive validity and reliability evidence are clear. But it is also evident, given the scientific complexity of obtaining such evidence, that there would be problems in this regard. The press of limited time to implement the new systems, coupled with lack of guidance on how to obtain defensible validity and reliability evidence, placed educators in the position of proceeding without appropriate safeguards in place. The professional standards of practice call for validity and reliability evidence before a program is made operational (AERA/APA/NCME, 1999). Without such persuasive evidence, the research community and professional vendors are obligated to mobilize quickly to address the need for this information. This research will probably require the combined efforts of both the special education community and testing and assessment professionals. The lack of persuasive technical data on the defensibility of alternate assessments at present suggests the need for great caution in implementing any high-stakes consequences for either individual or systemic accountability as a result of alternate assessments.

Challenges Faced by Teachers Administering Portfolio Assessment

Despite the fact that the intent is that an alternate assessment portfolio be assembled as much as possible with the input of the student, it is clear that the students for whom the portfolio assessment is appropriate (e.g., students with significant cognitive disabilities) may be limited in their ability to provide such input. As a result, the composition of each student's portfolio is likely to be highly reliant on the expertise and training of the student's teacher. Teacher background can impact student performance in two ways: teacher capacity in providing instruction covered in the assessment and teacher capability in assembling student portfolios. Either or both factors have a powerful impact on student performance.

Studies of the assessment of students with disabilities indicate that special educators often lack familiarity with the content and knowledge, or content standards, covered on assessments (DeStefano, Shriner, and Lloyd, 2001). Content coverage in a high-stakes assessment context can be a challenge for all teachers. However, it can be a particular challenge when the inclusion of students with disabilities, particularly those with significant disabilities, have had limited prior exposure to the general education curriculum.

According to research conducted elsewhere by Kleinert, et al., (1999) "the alternate portfolio process seems more focused on an assessment of the teacher than on the student." (p. 97) This study highlights the need for further analysis regarding the "extent to which teacher experience, scope, and recency of teacher training, or other salient teacher characteristics were related to reported adoption of instructional practices and teacher perceptions of the benefits of the alternate assessment to their students." (Kleinert, et. al, 1999, p. 97)

There does appear to be some evidence that teachers with greater experience, expertise and training are likely to produce a portfolio which receives a higher score than a teacher new to the process of producing an alternate assessment for the first time. Kleinert, et al., (2000) raised this question in their research: "to what extent did teacher (e.g., experience, amount of training) and instructional (amount of student involvement in the construction of the portfolio) variables predict the portfolio score?" Thompson et al., (2001) identify the issue of teacher training and experience regarding performance assessment as the key to improved results for teachers and students. Numerous authors have discussed the importance of teacher experience and training in portfolio use (Thurlow et al., 1998, Coutinho & Malouf, 1993, Harris & Curran, 1998).

Harris and Curran's (1998) study regarding the impact of knowledge, attitudes and concerns about portfolio assessment looked specifically at the impact on special educators. Their research findings indicate "if special educators are to use portfolios in ways that provide maximum benefits to their students, then they need to have greater knowledge about portfolios" (Harris & Curran, 1998, p. 92). According to Worthen (1993) "the classroom teacher is the gatekeeper of effective alternative assessment." (p. 447) Worthen (1993) further states: "to

a much greater degree than in traditional assessment, the quality of alternative assessments will be directly affected by how well teachers are prepared in the relevant assessment skills.” (p. 448)

In addition, teacher attitudes toward the use of portfolio assessment may be impacted by training and experience (Harris & Curran, 1998, Cheong, 1993). According to Harris and Curran (1998), “teachers who are trained and experienced in portfolio use have highly positive attitudes towards them” (p. 84). Given the current, and growing, critical shortage of qualified special educators (Donovan & Cross, 2002; McLaughlin, Artiles & Pullin, 2001), the extent of teacher expertise in both special education and alternate assessment will be a problem with growing implications.

Turner, Baldwin, Kleinert, and Kearns (2000), discuss the impact of teacher understanding of the scoring rubric and the resulting impact on student scores. According to Turner, et al., (2000), “understanding the scoring rubric may allow some teachers to represent quality indicators that are not actually apparent in the classroom” (p. 74). These authors articulated a possibility that teachers could inflate performance on a portfolio assessment. (Turner et al., 2000). This possibility raises significant concern regarding both validity and reliability issues arising from the fact that a portfolio assessment could be administered to the same student by two different teachers and result in entirely different scores. These two widely different scores could result from simple fundamental differences in the teachers' understanding of the requirements in the scoring rubric, as well as the teachers' familiarity with the individual student. All of these factors present considerable questions about the validity and reliability of inferences made about portfolio assessment.

Harris and Curran (1998) also articulate a number of “practical” problems affecting teachers using portfolio assessment. They identify these “practical problems as “the time involved, the cost, problems with planning portfolios, organizing and managing their contents, and selection of containers and storage” (Harris & Curran, 1998, p. 84; see also Kampfer, Horvath, Kleinert, and Kearns; Cheong, 1993). Turner et. al, (2000) offer an observation regarding the typical length of an alternate assessment when it is conducted in a portfolio format and the demand on teacher time. “As such, some teachers may not be willing to put forth the effort required to create a portfolio that accurately represents the student’s current program” (Turner, et.al, 2000, p. 74). States must recognize that support must be provided for educators to ensure that the “practical” problems do not negatively impact the portfolio score.

Educators at the ground level are instrumental in the success of alternate assessment programs. They must know how to identify potential candidates for alternate assessment, the content standards covered in the assessment and how to teach that content, how to address participation issues in IEP meetings, how to compile portfolios, and how to make appropriate judgments about student performance. They must find a way to do this when the consequences of alternate assessment are linked to both student and systemic accountability and perhaps as well their own individual accountability. They must also find ways to accommodate the time and intellectual demands associated with alternate assessment in their already busy days. And, as the critical shortage of

qualified special educators continues to grow, there will probably be fewer and fewer local educators who have even a rudimentary special education background (McLaughlin, Artiles & Pullin, 2001), independent of an understanding of the assessment issues discussed here.

Massachusetts' Implementation of an Alternate Assessment System: One State's Response

In response to national initiatives for education reform, many states passed their own reform legislation. A closer look at one state's efforts at alternative assessment, provides useful examples of the challenges educators face in the implementation of an alternate assessment program.

On June 18, 1993 the Massachusetts legislature enacted the Massachusetts Education Reform Act (MERA), which called for the creation of a statewide general curriculum in the major academic disciplines, school improvement plans and a new high-stakes assessment test tied to high school graduation (French, 1998). In response to federally imposed timelines, the Massachusetts State Board of Education began an ambitious implementation process for the MERA. A Five Year Master Plan organized five strategic goals which included eighty new initiatives. Among these initiatives was the development of the Massachusetts Curriculum Frameworks and the Massachusetts Comprehensive Assessment System (MCAS). Similar to other states statewide assessment systems, the MCAS is used for both systemic accountability (school and district performance indicators and potential state take-over of low performing schools or districts) and student accountability (individual student performance reports and high school graduation contingent upon acceptable MCAS performance). The MCAS is a large-scale, criterion-referenced testing system with provisions for accommodations for students with most disabilities.

For a student with disabilities, the IEP team is charged with determining whether the student 1) can take the standard MCAS under routine conditions, 2) can take the standard MCAS with accommodations, or 3) requires an alternate assessment. State guidelines instruct IEP teams in their decision-making based on the characteristics of a student's instructional program and local assessment (Mass. Dept. of Ed, 2002).

Massachusetts began the early stages of implementation of an alternate assessment system for students with significant disabilities in 1999. The state developed a portfolio-based assessment which was designed to measure student's knowledge of the key concepts and skills articulated by the general learning standards for all students set forth in the Massachusetts Curriculum Frameworks. This portfolio-based alternate assessment is known as the Massachusetts Comprehensive Assessment System – Alternate (MCAS-Alt). "The alternate assessment is intended for the very small number of students who are unable to participate in the standard MCAS due to the nature and severity of their disabilities" (Mass. Dept. of Ed, 2002, p. 16). For students with disabilities, "the purpose of the MCAS Alternate Assessment is to measure the achievement of these students on the Massachusetts Curriculum Framework learning standards in English Language Arts, Mathematics, Science and Technology/Engineering, and History and Social Science" (Mass. Dept. of Ed,

2000, p. 3).

The MCAS-Alt requires the collection of a body of evidence that may include student work samples, instructional data on the student, videotapes, and other supporting information linked to instruction in the subject being assessed. The training materials for educators provided by the Massachusetts Department of Education include a scoring guide which is intended “to help teachers and students prepare high-quality portfolio entries.” (Mass. Dept. of Ed, 2000, p. 23) According to the Massachusetts Department of Education, “the portfolio is developed over the course of the school year by the student, the student’s teacher, and other adults in the school or program who work with the student” (Mass. Dept. of Ed, 2002, p. 16).

The Massachusetts alternate assessment system has been described by one of the leading researchers on the testing of individuals with disabilities as “leading the way in the assessment and reporting of students with significant disabilities who require alternate assessments” (Thurlow, as quoted by Mass. Dept. of Ed., 2003). An examination of this system provides the opportunity to highlight some of the particular challenges confronting educators in implementing these reforms for students with significant disabilities. In terms of Maurer’s call for clarity, the goals of Massachusetts’ alternate assessment seem, on their face, to be clear. But, the question remains whether the assessment can meet the validity and reliability requirements regarding alignment of the “assessment content and the construct it is intended to measure” (AERA, APA, NCME, 1999, P. 11).

When the state of Massachusetts began to initiate its alternate assessment program in 1999, there were short timelines for implementation of the new assessments mandated by the federal government in the 1997 IDEA amendments. A system of assessment had to be developed and a large number of educators that had to be trained to administer the MCAS-Alt. Massachusetts field tested the MCAS Alternate Assessment during the 1999-2000 school year. During the 2000-2001 school year the alternate assessment was officially implemented for the first time, with the first portfolio assessments due at the beginning of May 2001.

Between October 2000 and January 2001, the Massachusetts Department of Education trained 3300 administrators and teachers in the implementation process of the MCAS-Alt. The deadlines of the federal mandates had a significant impact on the effectiveness of this training. According to Dan Wiener, Project Coordinator of the MCAS-ALT for the Massachusetts Department of Education, “it became clear that we needed to train teachers very intensively and give them much more time than we gave them, which we had every intention of doing but the law gave us such a short, brief, turnaround time” (Wiener, 2002a).

Additional challenges associated with the implementation of alternate assessment were concerned with how the evidence would be assessed and scored (Weiner, 2002b). The scoring rubric for the MCAS-Alt developed by its private testing contractor is used to review, evaluate and score student portfolios. Scorers examine each portfolio strand for evidence of the student’s performance in the following categories: completeness of materials submitted;

demonstration of the level of complexity at which the student addresses the learning standards in each content area; demonstration of the accuracy of the student's responses and performance on each product; evidence of the degree of independence the student demonstrated in performing each task or activity; and evidence of the student's ability to make decisions and/or self-evaluate as they engage in the task or activity (Mass. Dept. of Ed, 2002).

The scoring rubric is used to generate a numerical score for each portfolio strand and then the three scores of the three portfolio strands submitted in each content area are averaged in order to determine an overall score. The overall scores are translated into performance levels by the Massachusetts Department of Education in conjunction with its assessment contractor. The performance levels used to report student results in each content area in which the MCAS-Alt is administered include the three performance levels used in the standard MCAS (needs improvement, proficient, and advanced) as well as three additional areas (awareness, emerging, and progressing). A description of the performance levels for the MCAS Alt is as follows: awareness (student demonstrates very little understanding of learning standards), emerging (student demonstrates a rudimentary understanding of a limited number of learning standards and addresses the standards at substantially below grade level expectations), progressing (student demonstrates a partial understanding of some learning standards and address the standards at below grade level expectations), needs improvement (student demonstrates a partial understanding of the content area at grade level expectations), proficient (student demonstrates a solid understanding of the content area at grade level expectations), and advanced (student demonstrates a comprehensive and in-depth understanding of the content area at grade level expectations).

The scoring criteria for the rubric were determined with the assistance and feedback of hundreds of teachers who participated in the implementation of 1999-2000 field test. The scorers of the alternate assessments are recruited and trained by the Massachusetts Department of Education and its contractor. As the state itself confirmed, the difficulties of scoring alternate assessments represent a challenge “to use methods other than traditional testing to portray what a student has learned and to do this in a way that allows others who may not work directly with the student to interpret this evidence correctly” (Mass. Dept. of Ed, 2000, p. 23).

During the first year of implementation it became clear that “there were in some cases, different interpretations of the ways in which we told people to score” (D. Wiener, personal communication, Feb. 26, 2002). As a result, the state reevaluated the training system for scorers and made changes in the training plan for scorers for the next round of portfolio scoring.

The 2002 MCAS-Alt portfolios were scored during a three week scoring institute that was conducted in July 2002 during which 5300 MCAS-Alt portfolios were scored by 125 Massachusetts educators. Educators from across the state were recruited to participate in the scoring institute and preference was given to educators who could commit to the full three weeks of scoring. To prepare the scorers for the task of scoring the MCAS-Alt portfolios, scorers received a set of written scoring guidelines two to three weeks prior to the scoring institute. In

addition, the scorers participated in one and one-half days of training at the beginning of the scoring institute. Calibrated training strands were used to “qualify” scorers for the task of scoring the MCAS-Alt (Mass. Dept. of Ed, 2002). As a means of establishing reliability in the scoring, approximately 25% of the MCAS-Alt’s were scored by two different scorers. In addition, due to the significant consequences (award of a regular high school diploma) attached to the 10th grade score, all grade 10 MCAS-Alt’s were scored by two different scorers (Mass. Dept. of Ed , 2003, p. 2).

A similar scoring process was implemented in the 2003 administration. 5118 portfolios were scored by approximately 150 scorers during a three week scoring institute using a similar process as the 2002 scoring institute (Mass. Dept. of Ed, 2003).

According to the Massachusetts Department of Education, “It is anticipated that scores may be modest in the first few administrations of the MCAS Alternate Assessment, but scores are generally expected to improve...as educators become increasingly familiar with these requirements” (Mass. Dept. of Ed, 2002, p. 27). In fact, the data support this statement. Although changes in scoring make it impossible to clearly establish year-to-year trends, in each of the three years of administration of the MCAS Alt, approximately 1% of all the students in the state (about 6.5% of the students with disabilities in the state) participated in the alternate assessment. In 2001, 75% of the portfolios submitted scored in the lowest performance category “awareness”. In 2002, only 5% were scored at the “awareness” level, due in large part to a change in scoring. In 2003, only 3.5% were scored at the “awareness” level.

Changes in how the data was recorded from Year 1 (2001) to Year 2 (2002) are important to note. In the recording/categorization of the Year 1 data, those portfolios which were unable to be scored because there was insufficient evidence were included in the data for the awareness category. In the Year 2 data presentation, this data was separated out and an incomplete section was included in the data display. In Year 2, 44% of the portfolios were incomplete in at least one subject area. In the Year 2 results however, the combination of the incomplete data and the awareness data (49%) is lower than the Year 1 awareness data (75%) . Also of note, in Year 2, 34% of the portfolios scored in the progressing category an increase of 21% from Year 1. In Year 3 (2003) the percentage of portfolios which received incompletes dropped to 19% and the percentage of portfolios which scored in the progressing category increased to almost 65%, (D. Weiner, personal communication, 9/03).

The state reported in 2002 that it did include MCAS Alt data within its reports on the overall performance of all students in the state and all students with disabilities. Overall, on the Grade 10 MCAS, used to determine high school diploma awards, 14% of all students across the state failed and 45 % of students with disabilities failed (Mass. Dept. of Ed., 2002, August). Among the students participating in the alternate assessment, only 12 students across the state received a passing score (needs improvement or higher) on the Grade 10 level. (Mass. Dept of Ed., 2003). However, in 2003 the number of students that received a passing score increased to 26. "This number represents a dramatic increase over the previous two years" (Mass. Dept. of Ed., 2004).

Massachusetts is currently making an attempt to address requirements in the NCLB legislation regarding reporting of student assessment results. The state has made a plan for reporting the aggregated results in a manner which attempts to minimize the potential negative impact of the inclusion of student alternate assessment scores by assigning a point value system to the portfolios based on the scored performance level for each portfolio. The points would be assigned to the MCAS-Alt performance levels (0 points = portfolio not submitted, 25 points = incomplete, 50 points = awareness, 75 points = emerging, 100 points = progressing) in a similar manner as the regular MCAS (0 points – failing, 25 points – needs improvement, 50 points – proficient, and 100 points – advanced). The plan is for this reporting system to be implemented in the 2004 administration of the MCAS and MCAS-Alt.

In addition to challenges associated with scoring the MCAS Alt, there are also issues concerning content coverage for the assessment. In Massachusetts the alternate assessment is linked directly to the general education standards in the Massachusetts Curriculum Frameworks and is intended to assess student's mastery of skills, concepts and information regarding the general curriculum. Consistent with the state's regular assessment, the MCAS Alternate Assessment requires assessment in English Language Arts, Mathematics, History and Social Science, and Science and Technology/Engineering.

However, the MCAS Alternate Assessment does not include assessment in essential life areas or functional skills as has been the practice in some other states such as Maryland and Kentucky. According to Dan Wiener, Project Coordinator of the MCAS Alt at the Massachusetts Department of Education, "I think we're in the minority in that we haven't...but many access skills are embedded in the entry points to our Curriculum Frameworks" (personal communication, Feb. 26, 2002).

In response to the need to make the general curriculum accessible to all students, a resource guide was developed by the Massachusetts Department of Education which includes "instructional and assessment strategies [that] provide opportunities to teach students with disabilities the same standards as general education students, and to promote greater 'access to the general curriculum' for students with disabilities, as required by law" (Mass. Dept. of Ed, 2002).

The educator's manual describes four ways that students with disabilities can participate in the general curriculum. Those four areas are: (1) addressing the standard as written for the grade level of the student; (2) addressing the standard as written but using a different method of presentation and/or student response; (3) addressing the standard at lower levels of complexity and difficulty than grade-level peers, and (4) addressing the standard through social, communication, and motor "access skills" that are "incorporated and embedded in standards-based learning activities" (Mass. Dept. of Ed, 2002, p. 56).

Jacqueline Farmer Kearns, Project Director of the Interdisciplinary Human Development Institute at the University of Kentucky states that "access skills are a way that students with disabilities can participate in the general curriculum" (J. Farmer Kearns, personal communication, April 21, 2000). In the *2003 Educator's Manual for the MCAS Alternate Assessment* (Mass. Dept. of Ed,

2003), the state describes access skills in the following manner: “skills become ‘access skills’ when they are practiced as a natural part of instruction based on learning standards. When students practice their skills during daily academic instruction, they are participating in the general curriculum, though at a very basic level” (p. 57).

Administering an alternate assessment based on alignment with the general curriculum has added yet another layer of difficulty in the quest for education reform. It is well recognized that the federal mandate to adapt and align the general curriculum for all students including students with significant disabilities has presented a challenge for school districts across the country. A recent study of Massachusetts teachers of students with significant disabilities who participated in MCAS Alt elicited evidence from teachers that their students’ participation in the assessment process did cause teachers to pay attention to state curriculum frameworks they had previously ignored. These teachers also indicated the importance of the provision of appropriate and ongoing professional development activities at the state and building level which address the issues related to administering the MCAS-Alt with students with significant disabilities including assistance with curriculum alignment for this population. The study concludes that school districts should seek to use trainers/consultants who have experience with administering the MCAS-Alt and with aligning curriculum for students with significant disabilities. (Zatta, 2003)

In the past four years of administration of the MCAS Alt (one pilot year and three statewide administrations), it has become clear that the resources to assist teachers with the administration of an alternate assessment have increased but still have failed to adequately address the needs of students with significant cognitive disabilities and the educators who serve them. This is particularly true in the area of professional development. As Richard Elmore (2002) asserts, “the pedagogy of professional developers [must] be as consistent as possible with the pedagogy that they expect from educators. It has to involve professional developers who, through expert practice, can model what they expect of the people with whom they are working (p. 8).” Effective training efforts serve to increase capacity not only on an individual teacher-by-teacher basis but at the building and system level as well. Building capacity not only serves to ensure effective implementation but supports sustained reform as well.

Several variables related to professional development activities were found to impact the effectiveness of the administration of the MCAS-Alt. These variables included: teacher understanding, teacher willingness, commitment from school leadership and availability of resources. Developing understanding and willingness amongst the individuals responsible for the administration of the MCAS-Alt is important to the resulting student outcomes. The resources identified as having an impact on the administration of the MCAS-Alt include the availability of consultants experienced in the assessment system, peer support, sufficient time to implement the program, and adequate materials and equipment (Zatta, 2003).

Training in the specifics of the scoring guidelines of the alternate assessment has also been identified as important in terms of the potential impact on student

scores. Teachers in Massachusetts indicated that experience with the scoring rubric of the MCAS-Alt gave them a clearer understanding of the specific requirements. Those who had participated in pilot studies during the development of MCAS Alt and in scoring sessions for the assessment felt the most competent to effectively participate in the assessment system (Zatta, 2003). "Of course, as teachers also gain familiarity with portfolio management techniques, submission requirements, curriculum alignment, and instructional improvements, the scores of all students will rise" (Weiner, 2002b, p. 9). Training specifically targeted to the teachers of students with significant disabilities and experience with the scoring rubric were regarded by teachers as critical in providing them with the information needed to effectively administer the MCAS-Alt (Zatta, 2003).

In addition, the issue of training for scorers and the impact of training on the resulting student scores was also identified as an area of importance. Teachers questioned the reliability of their students' scores based on a comparison of the comments made by different scorers regarding similar portfolio evidence. The issues of scorer training must be carefully attended to in order to maximize inter-rater reliability. This issue is not unique to Massachusetts. A study conducted in Kentucky in 1999 also called for more research regarding the "development of performance-based measures for students with significant disabilities to meet the rigorous technical requirements of inter-rater scoring reliability" (Kleinert et al., 1999, p. 100).

The 2003 annual training for administrators responsible for the implementation of the MCAS-Alt in their respective schools underscored the importance of support from school leadership as well as an emphasis on training for teachers (Mass. Dept. of Ed, 2003). This shift in emphasis from previous yearly training focused exclusively on teachers may be indicative of the state's recognition of the importance of leadership issues in the alternate assessment program.

The Massachusetts alternate assessment system is but one approach to the challenges associated with including students with disabilities in education reform and accountability efforts. At this juncture, the state is only in the early stages of implementing its system. The evidence reported here point to further areas for future efforts to enhance the quality of alternate assessments and associated educational practices for students with significant disabilities.

Conclusion

The Congress set out a laudable series of goals when it required that students with disabilities be fully included in state and local standards-based education reform initiatives. It is clear that the intent of the federal and state legislation is to improve current practices within the entire education system. It is also clear that the current initiatives may not yet be fully and appropriately including the low incidence population of students with significant disabilities. In their zeal to call for a unified system of educational accountability and correct the problems of exclusion in the past, legislators and policymakers alike have not always recognized the individual and intensive needs of children with significant cognitive disabilities. Nor have they recognized the many unresolved issues associated with alternate assessment. As a result, significant further efforts are

needed to develop and refine the processes for assessing students with significant disabilities. These efforts must involve both educators and policy-makers at the ground level, as well as the private vendors who design and deliver assessment systems. Equally important, the research community faces considerable challenges in both assessing the effects of these assessments as well as offering scientifically-based solutions to the challenges associated with alternate assessment.

The goals of education reform are substantial and complex. It is no wonder that there are such daunting issues related to how to effectively achieve full participation for low incidence populations such as individuals with significant cognitive disabilities. Yet, at the same time, these students must not be overlooked. Now is the time to begin to consider how to better include and account for their abilities. As one disability advocate has noted, "we have moved from access to the schoolhouse to access to high expectations and access to the general curriculum" (Warlick, 2000, p. 11). The challenge ahead is to realize the goal of full and effective participation for students with significant disabilities.

References

- American Educational Research Association, American Psychological Association, and National Council for Measurement in Education (1999). *Standards for Educational and Psychological Testing*: Washington, D.C.: American Research Association.
- Board of Education v. Ambach, 458 N.Y.S. 2d 680 (1982).
- Brookhart v. Illinois State Board of Education, 697 F2d 179, 182 (7th Circuit, 1983).
- Carpenter, C. D., Ray, M., & Bloom, L. (1995). Portfolio assessment: Opportunities and challenges. *Intervention in School and Clinic*, 31(1), 34-41.
- Chapman v. California Department of Education, 229 F. Supp. 981 (N.D. Calif., 2002)
- Cheong, J. (1993). Portfolios: A window on student achievement. *Thrust for Educational Leadership*(November/December), 27-29.
- Choate, J., & Evans, S. (1992). Authentic assessment of special learners: Problem or promise? *Preventing School Failure*, 37(1), 6-9.
- Coutinho, M., & Malouf, D. (1993). Performance assessment and children with disabilities: Issues and possibilities. *Teaching Exceptional Children*, 25(4), 63-67.
- Debra P. v. Turlington, 644 F2d 397 (5th Cir. 1981), also F2d. (11th Cir. 1984).
- DeStefano, L, J. Shriner, and C. Lloyd (2001, Fall). Teacher decision-making in participation of students with disabilities in large-scale assessment. *Exceptional Children* 68:1, pp. 7-22.
- Donovan, M.S. & Cross, C. (eds.) (2002). *Minority Students in Gifted and Special Education*. Washington, D.C.: National Academy Press.
- Elliott, J. (1997). Invited commentary. *TASH*, 22(2), 104-106.
- Erickson, R., & Thurlow, M. (1997). *State special education outcomes, 1997: A report on state activities during educational reform*. Minneapolis: National Center for Educational Outcomes.
- French, D. (1998). The state's role in shaping a progressive vision of public education. *Phi Delta Kappan*(November), 185-195.
- Geenen, K., Thurlow, M., & Ysseldyke, J. (1995). *A disability perspective on five years of education*

- reform*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Goals 2000: Educate America Act, P.L. 103-227 (1994).
- Guy, B., Shin, H., Lee, S.-Y., & Thurlow, M. (1999). *State graduation requirements for students with and without disabilities* (Technical Report 24). Minneapolis: National Center for Educational Outcomes.
- Harris, M., & Curran, C. (1998). Knowledge, attitudes, and concerns about portfolio assessment: An exploratory study. *Teacher Education and Special Education*, 21(2), 83-94.
- Heaney, K. & D. Pullin (1998). Accommodations, Flags, and Other Dilemmas: Disability Rights and Admissions Testing. *Educational Assessment*, 5 (2), 71-93.
- Kampfer, S., L. Horvath, H. Kleinert, and J. Kearns (2001, Spring). Teachers' perceptions of one state's alternate assessment: Implications for practice and preparation. *Exceptional Children* 67:1, pp. 361-374.
- Kleinert, H., Haig, J., Kearns, J. F., & Kennedy, S. (2000). Alternate assessments: Lessons learned and roads to be taken. *Exceptional Children*, 67(1), 51-66.
- Kleinert, H., & Kearns, J. F. (1999). A validation study of the performance and learner outcomes of Kentucky alternate assessment for students with significant disabilities. *The Association for Persons with Severe Handicaps*, 24(2), 100-110.
- Koretz, D., McCaffrey, D., Klein, S., Bell, R., & Stecher, B. (1993). *The reliability of scores from the 1992 Vermont portfolio assessment program* (CSE Technical Report 355). Los Angeles: University of California.
- Maurer, R. (1996). *Designing alternative assessments for interdisciplinary curriculum in middle and secondary schools*. Needham, MA: Allyn & Bacon.
- Massachusetts Department of Education (2003). MCAS Alternate Assessment Summary of 2001 and 2002 Results. Malden, MA: The Commonwealth of Massachusetts.
- Massachusetts Department of Education (2004). MCAS Alternate Assessment (MCAS-Alt): Summary of 2003 State Results. Malden, MA: The Commonwealth of Massachusetts.
- Massachusetts Department of Education, (2003). MCAS Alternate Assessment 2004 Administrator's Training, Framingham, MA.
- Massachusetts Department of Education, (2002). 2003 MCAS Alternate Assessment Educators' Training, Brookline, MA..
- Massachusetts Department of Education (2002, August). Spring 2002 MCAS Tests: Summary of State Results. Malden, MA: The Commonwealth of Massachusetts.
- Massachusetts Department of Education (2002). Requirements for the Participation of Students with Disabilities in MCAS – Including Test Accommodations and Alternate Assessment, Malden, MA: The Commonwealth of Massachusetts.
- Massachusetts Department of Education (2001). Educator's Manual for the MCAS Alternate Assessment, Malden, MA: The Commonwealth of Massachusetts.
- The Massachusetts Education Reform Act of 1993: Research and Evaluation Mapping Project* (1999). Amherst, MA: University of Massachusetts Amherst School of Education.
- McDonnell, L., McLaughlin, M., & Morison, P. (1997). *Educating one and all*. Washington, D.C.: National Academy Press.
- McLaughlin, M., Artiles, A. & Pullin, D. (2001). Challenges for transformation of special education in the 21st century: Rethinking culture in school reform. *Journal of Special Education Leadership*.
- Miller, M. D., & Legg, S. M. (1993). Alternative assessment in a high-stakes environment. *Educational Measurement: Issues and Practice*(2), 9-15.
- Moss, P. (1992). Shifting conceptions of validity in educational measurement: Implications for

- performance assessment. *Review of Educational Research*, 62(3), 229-258.
- National Center on Educational Outcomes. (1999). *1999 state special education outcomes: A report on state activities at the end of the century*. Minneapolis, MN: University of Minnesota.
- National Center on Educational Outcomes. (2000). *IEP's and standards: What they say for students with disabilities*. Minneapolis, MN: University of Minnesota.
- National Center on Educational Outcomes, (2000). *Special topic area: Accountability for students with disabilities*. Minneapolis, MN: University of Minnesota.
- National Commission on Excellence in Education (1983). *A nation at risk: The imperative for education reform*. Washington, D.C.: U.S. Government Printing Office.
- No Child Left Behind Act of 2001, P.L. 107-110.
- Nolet, V. (1992). Classroom-based measurement and portfolio assessment. *Diagnostique*, 18(1), 5-24.
- Olsen, K. (1998). *What principles are driving development of state alternate assessments?:* Mid-South Regional Resource Center.
- Quenemoen, R., Thompson, S. and Thurlow, M. (2003). Measuring academic achievement of students with significant cognitive disabilities: Building understanding of alternate assessment scoring criteria (Synthesis Report 50). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Rouse, M., Shriner, J., & Danielson, L. (2000). National assessment and special education in the United States and England and Wales: Towards a common system for all? In M. McLaughlin & M. Rouse (Eds.), *Special Education and School Reform in the United States and Britain* (pp. 67-97). London and New York: Routledge.
- Sailor, W. (1997). Invited Commentary. *TASH*, 22(2), 102-114.
- Sebba, J., Thurlow, M., & Goertz, M. (2000). Educational accountability and Students with Disabilities in the United States and Canada. In M. McLaughlin & M. Rouse (Eds.), *Special Education and School Reform in the United States and Britain*. New York: Routledge.
- Shepard, L. (1992). *Will national tests improve student learning?* (CSE Technical Report 342). Los Angeles: University of California.
- Standards for Educational and Psychological Testing*(1999). Washinton, D.C.: American Educational Research Association.
- Thompson, S., Quenemoen, R., Thurlow, M., & Ysseldyke, J. (2001). *Alternate Assessments for Students with Disabilities*. Thousand Oaks, California: Corwin Press, Inc.
- Turner, M., Baldwin, L., Kleinert, H., & Kearns, J. F. (2000). The relation of statewide alternate assessment for students with severe disabilities to other measures of instructional effectiveness. *The Journal of Special Education*, 34(2), 69-76.
- Warlick, K. (2000, June 23-24, 2000). *Proceedings Report*. Paper presented at the Alternate Assessment Forum: Connecting into a Whole, Salt Lake City, Utah.
- Warlick, K., & Olsen, K. (1998). *Who to assess? State practices in determining eligibility for alternate assessment.*: Mid-South Regional Resource Center.
- Weiner, D. (2002a, February). Personal interview with M. Zatta.
- Wiener, D. (2002b). *Massachusetts: One state's approach to setting performance levels on the alternate assessment*(Synthesis Report 48). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved July 14, 2003, from the World Wide Web: <http://education.umn.edu/NCEO/OnlinePubs/Synthesis48.html>
- Worthen, B. (1993). Critical issues that will determine the future of alternative assessment. *Phi Delta Kappan*, 444-456.

- Ysseldyke, J., Olsen, K., & Thurlow, M. (1997). *Issues and considerations in alternate assessments* (Synthesis Report No. 27). Minneapolis, MN: University of Minnesota: National Center on Educational Outcomes.
- Ysseldyke, J., & Olson, K. (1997). *Putting alternate assessments into practice: What to measure and possible sources of data*. Minneapolis: National Center for Educational Outcomes.
- Ysseldyke, J., & Thurlow, M. (1999). *Assessment and accountability for student with disabilities: 1999 status update and emergin issues*. Minneapolis: National Center for Educational Outcomes.
- Ysseldyke, J., Thurlow, M., & Geenen, K. (1994). *Educational accountability for students with disabilities*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Ysseldyke, J., Thurlow, M., Kozleski, E., & Reachly, D. (1998). *Accountability for the results of educating students with disabilities: Assessment provisions of the 1997 Ammendments to the Individuals wiht Disabilities Education Act*. Minneapolis: National Center for Educational Outcomes.
- Ysseldyke, J., Thurlow, M., & Shriner, J. (1994). *Students with disabilities & educational standards: Recommendations for policy & practice*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Zatta, M. (2003). *Is there a relationship between teacher experience and training and student scores on the MCAS Alernate Assessment?* (Unpublished doctoral dissertation, Boston College, Chestnut Hill, Massachusetts).

About the Authors

Mary Zatta

Perkins School for the Blind
Email: mary.Zatta@Perkins.org

Mary Zatta received her Ph.D. from Boston College. She is administrator in the Deafblind Program at Perkins School for the Blind in Watertown, Massachusetts. At the Perkins School, she is responsible for educational and residential programming for deafblind adolescents. In addition to her work at Perkins, she has served as an international consultant in several nations on issues related to the instruction of deafblind children and adolescents. In addition, she is an adjunct faculty member at the University of Massachussetts-Boston / Center for Social Development and Education.

Diana Pullin

Lynch School of Education
School of Law
Boston College
Email: pullin@bc.edu

Diana Pullin is Professor of Education Law and Public Policy at the Lynch School of Education and the School of Law at Boston College. She holds a law degree and a Ph.D. from The University of Iowa. She has published extensively in the area of education law and public policy and has served as a consultant to numerous professional associations, research centers, advocacy groups, attorneys, and education officials on issues concerning law, testing, and disability.

The World Wide Web address for the *Education Policy Analysis Archives* is
epaa.asu.edu

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, [Gene V Glass, glass@asu.edu](mailto:glass@asu.edu) or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu.

EPAA Editorial Board

[Michael W. Apple](#)
University of Wisconsin

[Greg Camilli](#)
Rutgers University

[Sherman Dorn](#)
University of South Florida

[Gustavo E. Fischman](#)
Arizona State University

[Thomas F. Green](#)
Syracuse University

[Craig B. Howley](#)
Appalachia Educational Laboratory

[Patricia Fey Jarvis](#)
Seattle, Washington

[Benjamin Levin](#)
University of Manitoba

[Les McLean](#)
University of Toronto

[Michele Moses](#)
Arizona State University

[Anthony G. Rud Jr.](#)
Purdue University

[Michael Scriven](#)
University of Auckland

[Robert E. Stake](#)
University of Illinois—UC

[Terrence G. Wiley](#)
Arizona State University

[David C. Berliner](#)
Arizona State University

[Linda Darling-Hammond](#)
Stanford University

[Mark E. Fetler](#)
California Commission on Teacher
Credentialing

[Richard Garlikov](#)
Birmingham, Alabama

[Aimee Howley](#)
Ohio University

[William Hunter](#)
University of Ontario Institute of
Technology

[Daniel Kallós](#)
Umeå University

[Thomas Mauhs-Pugh](#)
Green Mountain College

[Heinrich Mintrop](#)
University of California, Los Angeles

[Gary Orfield](#)
Harvard University

[Jay Paredes Scribner](#)
University of Missouri

[Lorrie A. Shepard](#)
University of Colorado, Boulder

[Kevin Welner](#)
University of Colorado, Boulder

[John Willinsky](#)
University of British Columbia

EPAA Spanish and Portuguese Language Editorial Board

Associate Editors for Spanish & Portuguese

Gustavo E. Fischman
Arizona State University
fischman@asu.edu

Pablo Gentili
Laboratório de Políticas Públicas
Universidade do Estado do Rio de Janeiro
pablo@lpp-uerj.net

Founding Associate Editor for Spanish Language (1998-2003)
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia
Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Universidad Autónoma de Puebla
rkent@puebla.megared.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidade Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
American Institutes for
Research-Brazil (AIRBrasil)
simon@sman.com.br

[Jurjo Torres Santomé](#) (Spain)
Universidad de A Coruña
jurjo@udc.es

[Carlos Alberto Torres](#) (U.S.A.)
University of California, Los Angeles
torres@gseis UCLA.edu

EPAA is published by the Education Policy Studies
Laboratory, Arizona State University